

多重の読みを持つ宣命コーパスの構築

呉 寧真 (国立国語研究所)

池田 幸恵 (中央大学)

須永 哲也 (昭和女子大学)

小木曾 智信 (国立国語研究所)

国立国語研究所の『日本語歴史コーパス』の一部として、8世紀に書かれた『続日本紀』宣命を収録する「奈良時代編Ⅱ宣命」を追加・公開した。上代の公的文書は一般に漢文体で書かれ、原文の解読が難しい。一方、宣命は和文体で、助詞・助動詞等が一字一音の万葉仮名を用いて書かれており、音読が困難な時代のテキストの読解と発音を知る重要な日本語資料である。コーパス化にあたっては、漢字と万葉仮名によって表記されている原文を、独自の漢字仮名交じりテキストに変換して形態論情報を付与していた。また、他の資料では一般的に音読する漢語を訓読している場合など、表記と単語情報が単純に対応しない問題を解決するために、同一箇所に複数の形態論情報を付与した。

Construction of the corpus of "Senmyō" with multiple readings

Neisin Go (National Institute for Japanese Language and Linguistics)

Yukie Ikeda (Chuo University)

Tetsuya Sunaga (Showa Women's University)

Toshinobu Ogiso (National Institute for Japanese Language and Linguistics)

As a part of the Corpus of Historical Japanese, the National Institute for Japanese Language and Linguistics created a morphologically annotated corpus of the *senmyō* in the "Shoku Nihongi," written in the 8th century. During this period official Japanese texts were usually written in Classical Chinese, but *senmyō* were written in native Japanese. Because function words such as particles and adverbs are written in these texts in *man'yōgana*, which indicate pronunciations, *senmyō* constitute a very important material of the Japanese language providing the reading and pronunciation of texts from a period for which transliteration is difficult. In making this corpus, we prepared both the original Chinese text and the translated Japanese text and made it possible to refer to both texts at the same time. In order to resolve difficulties where the notation and word information do not simply correspond to each other (for example, in cases of uncommon readings for Chinese characters), we provided multiple entries in the morphological information assigned to the item in question.

1. はじめに

国立国語研究所では、古代語から近代語までの資料を網羅する『日本語歴史コーパス』の構築に取り組んでいる。2020年3月、「奈良時代編Ⅱ宣命」として新たに『続日本紀』に残されている62編の宣命、約2万語のデータを追加し、公開した。宣命の原文は漢字と万葉仮名で表記され、漢文の語彙・語順を用いる場合があることから、表記と読みの対応が難解である。

そこで本研究では、表記と読みの対応がとれない問題を解決するために、同一箇所に複数の形態論情報を付与する手法を提案する。また、この手法が漢語・漢文を用いる他の資料にも応用可能であることを述べる。

2. 続日本紀宣命とは

『続日本紀』とは、奈良時代の歴史書である。『続日本紀』に収録された「宣命」の部分は、口頭で宣布された天皇の詔であるため、和文体で記されている。宣命では、体言や用言の語幹などは漢字を用いて、大字で書かれており、助詞・助動詞・用言の活用語尾などは一字一音の万葉仮名を用いて、小字で書かれている。このような表記体は「宣命書き」と呼ばれる。公的文書が漢文体で書かれることが主流であった上代において、一字一音の万葉仮名を用いる宣命は、当時の和文体の資料として価値の高い日本語資料である。

例えば、図1の左頁2行目では、「高天原由天降坐」に、万葉仮名「由(ゆ)」を動作の起点を表す格助詞として使用している。これは中古以降には見られない語である。

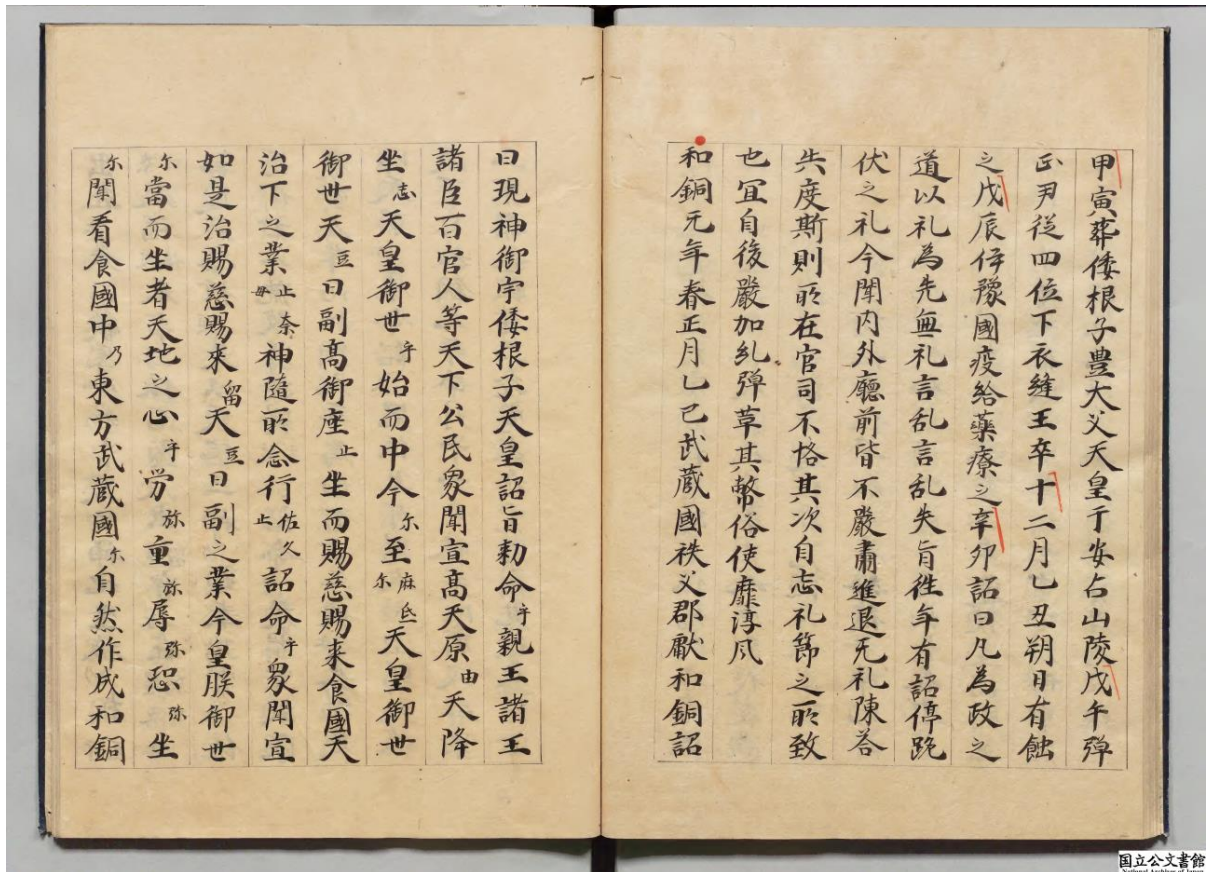


図1『続日本紀』の本文（右頁）と宣命部分（左頁）[5]

3. 宣命コーパスの構築

宣命の研究においては、原文の漢字・万葉仮名が参照されることが多く、コーパス化にあたっては、原文が利用できるような設計することが望まれた。一方で原文そのままの表記ではコーパスの利用者が一見して意味を取りにくいという問題があり、漢字仮名交じりに変換したテキストも必要となった。

本コーパスは、漢字と万葉仮名によって表記されている原文を、独自の漢字仮名交じりテキストに変換して形態論情報を付与している。これは漢字が連続する文字列や訓読を前提とした漢文的な語順のままでは、形態素解析が困難であること、また『日本語歴史コーパス』の他のサブコーパスの検索結果と並べて表示した際に視認性が下がることなどの理由による。

以下、校訂の際に依拠した北川和秀(1982)『続日本紀宣命 校本・総索引』を「底本」、本コーパス構築に当たって作成したテキストを「校訂本文」、底本の漢字・万葉仮名表記文を「原文」と呼ぶ。

【原文】

現御神止大八嶋国所知天皇大命良麻止詔大命乎集侍皇子等王等百官人等天下公民諸聞食止詔

【校訂本文】

現御神と大八嶋国所知す天皇が大命らまと詔りたまふ大命を、集り侍る皇子等王等百官の人等天の下の公民諸聞き食へと詔りたまふ。

校訂本文の作成にあたり、具体的には、①から⑦までの作業を行った。

①万葉仮名の処理

原文が万葉仮名である場合、平仮名に変換する。〔 〕内は原文表記、以下同。）

【例】掛けまくも畏き天皇〔掛母畏支天皇〕

【例】中今に至るまでに〔中今至麻弓尔〕

②補読の処理

底本では、原文に表示されない送り仮名・助詞・助動詞・敬語などを振り仮名で補読している。それらをも本コーパスの校訂本文は平仮名で表示することとした。

【例】官々に仕_へ奉_る韓人部一二人に〔官々仕奉韓人部一二人尔〕

【例】自然に作成_{れる}和銅〔自然作成和銅〕

【例】諸聞き食へと詔りたまふ〔諸聞食止詔〕

③返読の処理

原文が漢文的な語順で、訓読する際には字順を転倒させて読む場合、校訂本文はすべて返読処理をした。

【例】位有る人等に位一階給ふ〔有位人等給位一階〕

【例】神ながら所念し行す〔隨神所念行〕

④漢文的な語順による衍字の処理

漢文を和文に読み下す場合、一部の文字が衍字になることがある。特に次の【例】のような「動詞連体形＋格助詞＋名詞」の構造が散見される。この場合、助詞「之」は衍字（不読字）であり、校訂本文では表記しない。

【例】諂ひ欺く心無く忠に赤き誠を以て〔無諂欺之_レ心以忠赤_レ誠〕

【例】天の授けぬを得て〔天乃不授_所乎得天〕

⑤漢文的な語順における万葉仮名の衍字の処理

漢文を和文に読み下す場合、原文の「打消の「不」＋動詞」を「動詞＋打消の助動詞」に返読するが、打消の助動詞に当たる万葉仮名が同時に存在する構造が散見される。この場合、校訂本文では打消にあたる要素は一回しか表記しない。

【例】進みも知らに退きも知らに〔進母_不知_不退母_不知_不〕

【例】荷重きは堪へじかと〔荷重波_不堪_自加止〕

⑥踊り字の処理

原文には「々」「ゝ」の2種類の踊り字がある。

「々」については、漢字一字の繰り返しで、かつ前接文字と合わせて一語をなす場合（「人々」など）はそのまま「々」を用いて、それ以外は適宜繰り返し返されると想定される文字に置き換えた。なお、万葉仮名は平仮名に変換した。

【例】人々好からぬ謀を〔人々_不好謀乎〕

【例】御称称りて緩ひ怠る事無く〔御称々_而緩怠事無久〕

【例】天皇が御世御世聞し看し来る〔天皇御世々々聞看来〕

【例】よさし奉りしまにまに〔与佐斯奉志麻尔々々〕

「ゝ」は4例ある。これについてはすべて前接する万葉仮名に従って平仮名に変換した。

【例】思ほしつゝ大坐坐す〔思保之_ツゝ大坐坐〕

【例】所念しし位となも〔所念之_ゝ位止奈毛〕

⑦句読点の処理

底本では、漢文には句点・読点の双方があり、和文には句点だけが使用されている。本コーパスの校訂本文では、漢文では底本の句読点をそのまま使用し、和文では句の切れ目になる助詞、接続詞、動詞連用形などの直後に読点を付与した。句

点の付与は底本に従った。

【例】諸国々郡司、加位一階。〔諸国々郡司、加位一階。〕

【例】加以、元来風の病に苦びつつ、身体安らず。〔加以元来風病尔苦都々身体不安。〕

【例】受け賜り歡び、受け賜り貴び。〔受賜利歡受賜利貴〕

4. 問題の所在

3の手順により作成した校訂本文に対し、形態素解析辞書「UniDic」により形態素解析を行うが、宣命は和文体でありながら、漢文的な語順で表記される部分や漢文を含むため、形態素解析にあたって独自の問題が生じた。例えば、「おおやしましらしめす」という語を見てみよう。国立国語研究所のコーパスで用いられる「短単位」という言語単位では、「おおやしま」と「しらしめす」はそれぞれ1語（1短単位）となる。校訂本文では、「おおやしましらしめす」に対して、次の(1)と(2)の2通りの表記が見られ、このうち(2)においては、単語境界と表記との対応に問題が生じる。

(1) 大八洲所知す

(2) 御大八洲す

(1)はそれぞれ「大八洲」と「所知す」が対応するが、(2)は「しらしめす（御す）」が「おおやしま」を挟む形になり、表記上、「しらしめす」の形態論情報の付与が不可能であった。

このような状況は宣命だけではなく、例えば『今昔物語集』のような、漢文を多用する資料にも見られた。『日本語歴史コーパス』の従来のコーパスでは、表記上分割できない語については、次の3種類の手法で処理していた。

①例外的な短単位認定

僅少例に限り、例外として容認し、1短単位として認める。例えば、「以来（読み：このかた）」は、本来短単位の認定基準では、「こ | の | かた」と3短単位に分割する例であるが、表記上分割できないため、例外として1短単位とした。

②読みに合わせたテキストの校訂

補読・返読などによって、漢文的な語順で表記される部分を展開し、テキストを校訂する。例えば、「于今（読み：いまに）」は、「いま | に」と分割する例であり、読みの順に対応させるため、本文を「今に」に校訂した。

③表記に合わせた形態論情報の付与

底本で訓読みされていても音読みする。例えば、「二三年（読み：ふたとせみとせ）」では、訓読みは採用せず、「にさん | ねん」の音読みの形態論情報を付与した。

宣命も、①から③の手法で処理すれば、形態論

情報の付与は可能であるが、各々に課題が残る。

まず①については、前述の「おおよしましらしめす」のような短単位に対応せず、例外的な処理を要する語は、宣命の約2万語の中で1100語を超す。そのためそれらをすべて例外と扱い一語として認定すると、他のサブコーパスと統一がとれず、縦断的な検索ができなくなる。

また、②についての課題には次のようなことがある。例えば「御大八洲す」を返読し、テキストを校訂すれば、「大八洲(おおよしま) | 御す(しらしめす)」と形態論情報を付けることが可能である。ただし、宣命には短い文字列に多くの短単位が対応する例が多く、そのすべてに校訂を施すと、宣命の表記を保つことができない。宣命の原文に見られない表記をコーパス構築のために用いることになる。例えば、僧尼を統率する官「僧綱」の読みは「ほうしのつかさ」である。この読みに対し形態論情報を付与するために、テキストを「法師の司」と校訂すると、「僧綱」の表記が失われるばかりか、上代に僧尼を統率する官を意味する語の表記として「法師の司」が見られないことから、宣命の研究者にとって違和感があることになり、また、一般の使用者にも誤解が生じる可能性がある。

③についても課題がある。宣命の底本では訓読みされているが、地名と数字の読みは、『日本語歴史コーパス』では一貫して音読みすることが多く、コーパス全体としては音読みに統一すべきところである。例えば、「豊前」を「とよくに | の | みち | の | くち | の」の6単位に読むのではなく、「ぶぜん」と1単位にし、「十 | 五 | 日」を「とを | か | あまり | 一 | つ | か | の | ひ」の7単位に読むのではなく、「じゅう | ご | にち」と3単位にすると、形態論情報を付与することもでき、他のサブコーパスと統一させることもできる。しかし、地名や数詞に限って音読みを採用する場合、②の「僧綱」についても「そうごう」の音読みを採用すべきか否かなど、同一サブコーパス内での統一的な付与方針を定めるのが困難になる。

本研究では、上記のような表記と形態論情報の対応がとれない問題を解決するために、同一箇所にも複数の形態論情報を付与する手法を提案する。

5. 形態論情報の多重化

形態論情報の多重化とは、本文の同一箇所に複数の読み・形態論情報を付与できるようにした機能(小木曾 2016)によるもので、本来は掛詞や洒落などに対応するための機能であり、これまでに『日本語歴史コーパス』の「江戸時代編 I 洒落本」(村山ほか 2017)、「和歌集編」(松崎ほか 2019)、

「江戸時代編 III 近松浄瑠璃」(片山ほか 2020)で用いられている。例えば、名詞の「菊」に動詞の「聞く」が掛かっている場合、同じ箇所に複数の形態論情報を付与することができる。その場合、『日本語歴史コーパス』の規定では後の文脈につづくほうを「主本文」とし、もう一方を「副本文」とする。この例では「菊」が「主本文」であり、それに掛かる副次的な意味の「聞く」は「副本文」である。

- (1) なにしおへば長月ごとに君がためかきね
の菊は匂へとぞ思(後撰和歌集)

校訂本文	菊
主本文	語彙素：菊 品詞：名詞-普通名詞-一般
副本文	語彙素：聞く 品詞：動詞-一般

次の(2)(3)のように、副本文の単位の境界は必ずしも主本文と一致しない。しかし、本機能では下図のように副本文の形態論情報を必要に応じて範囲を変更して付与することができる(片山 2020b)。

- (2) 馬も太鼓をうつくしき踊り浴衣の上から
下まで、色めき喜び賑はへり(近松浄瑠璃)

校訂本文	う	つ	く	し	き
主本文	語彙素：美しい 品詞：形容詞-一般/文語形容詞-シク/ 連体形-一般				
副本文	語彙素：打つ 品詞：動詞- 一般/文語四 段-タ行/連体 形-一般				

- (3) いざ急がんとちよこ / \ 走り、とはかは
口にぞ着きにける(近松浄瑠璃)

校訂本文	と	は	か	は	口
主本文	—		語彙素：川口 品詞：名詞-普通名詞- 一般		
副本文	語彙素：とわか 品詞：副詞				

この機能が実装される以前のサブコーパスは、一つの文字列に対して一つの形態論情報しか付与できず、表記と形態論情報が対応できない場合、前節①から③の手法で処理するしかなかった。

文字列に対して付与される形態論情報の一つの情報の「層」と考えると、形態論情報の多重化を行っていないサブコーパスには、一つの「層」しかなかったのである。それに対して、形態論情報の多重化を行う場合、一つの文字列に複数の「層」を持たせることができる。5節の最初に挙げた「菊」と「聞く」の例では、文字列「菊」に対して、第一層に主たる意味の「菊」の形態論情報を付与し、第二層に「聞く」の形態論情報を付与した。この例では、第一層に付与した形態論情報が「主本文」、第二層の情報が「副本文」となる。

同一箇所にも3種類以上の語が掛けられている場合、第三層、第四層など層を増やして対応することが可能である。

『日本語歴史コーパス』の規定上、第一層の形態論情報は、表記と形態論情報の単語境界の一致が求められる。それに対し、第二層では文字列の範囲にかかわらず自由に形態論情報を付与することができる。この機能を活用し、第一層には表記に対応した音読みの語の形態論情報を付与し、表記との対応がとれない訓読みの形態論情報を第二層に付与した。例えば、表記「僧綱」の第一層には音読み「そうごう」を付与し、第二層には「ほうし | の | つかさ」の3短単位を付与した。

宣命研究においては、漢字表記語の読みについて、音読みよりも訓読みを優先させる方針がある。そのため、本コーパスでは第二層に形態論情報がある場合、それを主本文にすると設定した。多重化を用いて2層の情報を付与する手法はこれまで多重化を行ってきたサブコーパスと同じであるが、本コーパスでは第一層と第二層にある形態論情報の主副が交換されている点に大きな特徴がある。

本コーパスでは「二文字以上の漢字熟語と思われるもの」「地名」「数詞・助数詞」に複数の形態論情報を付与することとした。宣命の訓読は必ずしも一般的な読み方ではなく、宣命の研究者以外では、訓読を思いつかない場合が多いと想定される。そこで、コーパス検索者にとっての利便性を重視し、副本文は一般的な読み(音読み)とした。それによって検索者は比較的分かりやすい音読から検索できる。

【漢字熟語】

	校訂 本文	嫡子
第一層	副本文	ちやくし
第二層	主本文	むかいめ ばら の みこ

【地名】

	校訂 本文	越前
第一層	副本文	えちぜん
第二層	主本文	こし の みち の くち

【数詞・助数詞】

	校訂 本文	十	五	日
第一層	副本文	じゅう	ご	にち
第二層	主本文	とを か あまり いつ か の ひ		

漢字熟語の場合、副本文を付与するにあたって、以下の手順をとった。

【1】『日本国語大辞典 第二版』(以下『日本国語大辞典』)での項目確認

『日本国語大辞典』に表記が該当する項目があれば、たとえ初出が上代ではなくても音読みを付与した。文字列として出現したことを重視し、検索できるように、時代を考慮せず読みを与えた。

【2】『日本国語大辞典』での意味確認

同じ表記に二つ以上の語形がある場合は意味の異同があるかどうかを確認した。例えば、表記「行事」には、ギョウジとコウジの2語形があり、ギョウジは、「恒例として事を執り行なうこと。また、その事柄。催し事。「年中行事」の意味である。コウジは、「事を行なうこと。また、その人やその物事。」と記載されている。宣命の当該用例の文脈では、「催し事」の意味ではなく、「物事」の意味のため、語形「コウジ」を採用した。『日本国語大辞典』によると、コウジの初出は『史記抄』(1477年)と遅いが、本コーパスでは時代については考慮しない。

【3】その他

上記【2】のように、同じ表記に複数の読みが想定され、かつそれらに意味の差が認められない場合、初出が早い語形を採用した。それでも

なお決めがたい場合、『日本国語大辞典』の用例に、『続日本紀』や漢籍を引用している語形を採用した。

一方、第二層を活用することで、4節で挙げた「御大八洲す」のように、表記と語の順序が一致しない例についてもテキストを変更することなく形態論情報を付与することが可能となる。第二層では表記にかかわらず自由に形態論情報を付与することができるため、「御大八洲す」という文字列全体に対して「オオヤシマ」「知らしめす」の形態論情報を付与することができる。

ただし、「御大八洲」は「嫡子」や「越前」などと異なり、この文字列を音読みすることが考えられないため、第一層の形態論情報が空欄になる。そこで、ここでも主本文・副本文を交換し、第二層の形態論情報を主本文に設定した。本来は複数の形態論情報を付与するための機能を応用することで、テキストの校訂を最小限にとどめて形態論情報を付与することが可能となった。

	校訂本文	御大八洲す	
第一層	副本文	—	
第二層	主本文	語彙素：オオヤシマ 品詞：名詞-固有名詞-地名-一般	語彙素：知らしめす 品詞：動詞-一般

従って、宣命は既存のコーパスの多重化の目的・方法と異なり、二重の情報がある場合、主たる意味の形態論情報は、第二層にあるため、主本文と副本文の交換が必要である。また、「御大八洲す」のような例の場合は、実際に複数の意味があるのではなく、同一箇所複数の形態論情報の付与を行うことで、それを音読みするような例外的な読みを作ることを回避した。

以上のように、形態論情報の多重化の機能を用いることで、本コーパスは、宣命原文の表記を最大限に保ちながら、それらに形態論情報を対応させ、『日本語歴史コーパス』における他のサブコーパスとの統一化を図り、縦断的な検索を可能とした。

6. 長単位との連携

本コーパスには、言語的特徴の解明を目的とする検索のために、長単位の情報も付与している。長単位は短単位を基に作ってあるため、長単位の情報を持つ他のサブコーパスでは、第一層の形態

論情報を用いているが、本コーパスの第一層は、表1のように、主本文と副本文が混在している状況であり（色付けが主本文）、単一の層から長単位を構築する従来の方法では長単位情報がつかない箇所が生じる。そのため、第二層がある場合は第二層、ない場合だけ第一層と、多重情報を組み立て、主本文の情報として統一することを工夫した。

表1 情報が混在している宣命の層

校訂本文	今	御宇し	つる	天皇
第一層	いま	—	つる	てんのう
第二層	—	あめのした しらしめし	—	すめら みこと

表1の場合、「今」と「つる」の部分には第二層に情報がないため、従来通り、第一層にある情報を主本文とする。一方、第一層に情報がない「御宇し」は、主本文と副本文が交換されたため、第二層が主本文である。また、第一層に音読み、第二層に訓読みの情報が付与されている「天皇」も、訓読みのある第二層が主本文である。主本文の情報を第一層と第二層から組み上げた結果、「今御宇しつる天皇」の読みは「いまあめのしたしらしめしつるすめらみこと」になり、宣命の原文表記と訓読みを保つことができ、短単位と長単位の情報の一致もできた。

また、例えば、原文表記「嫡子」に対する訓読「むかいめばらのみこ」は助詞「の」が介在するため、助詞が介在すると分割するという長単位の規定に従うと、3長単位に分割される。しかし、本コーパスは主本文として原文表記の「嫡子」を分割できない単位としての情報を用いたため、表記を優先させ1長単位とした。

	校訂本文	嫡子
第一層	副本文	ちやくし
第二層	主本文	むかいめばらのみこ

従って、表記を重視し、表記の違いによって、同じ意味を表す同じ語の組み合わせでも、1長単位でとらえる場合ととらえない場合がある。例えば下記の例のように、「おおやしましらしめす」は表記によって、1長単位ととらえる場合と2長単位ととらえる場合があり、検索を行う際には注意が必要である。

|大八洲|所知す| (おおやしま|しらしめす)
|御大八洲す| (おおやしま=しらしめす)

検索対象 ^① 設定を隠す

検索対象を選択 検索対象をクリア

全て

検索動作 設定を隠す

文脈中の区切り記号 ^① | ▼ 文脈中の文区切り記号 ^① # ▼ 前後文脈の語数 ^① 20 ▼

副本文 ^① 副本文を検索対象に含む ▼ 共起条件の範囲 ^① 文境界をまたがない ▼

ダウンロード

副本文を検索対象に含む

副本文を検索対象に含まない

列の表示 ^① 設定を隠す

コーパス情報

時代名 サブコーパス名 サンプル ID 開始位置 連番 コア 層 層内連番 主本文 多重化種別

図2 副本文の選択

7. 検索と集計

本コーパスはこれまでの『日本語歴史コーパス』のサブコーパスと同様、検索アプリケーション「中納言」で検索することができる。本コーパスは部分的に形態論情報が多重になっていることで、検索対象に副本文を含むかどうか重要となる。検索する際、図2のように、「中納言」では、「副本文を検索対象に含む／副本文を検索対象に含まない」を選択することができる。「副本文を検索対象に含む」を選択すると、直下部の「列の表示」において「層」「主本文」「多重化情報」のチェックボックスにチェックが入る。検索結果の「主本文」列に「1」と表示される場合、検索結果は主本文のものであり、「0」と表示される場合は副本文のものである。

この設定による検索結果をダウンロードすると、主本文だけ、または主本文・副本文両方を含むデータをダウンロードできる。「主本文」列にある「1」「0」から、EXCEL等のフィルター機能を用いて、並べ替えや絞り込みができる。

また、『日本語歴史コーパス』のホームページには、語数表と語彙表を提供している（バージョン2020.03）。ここにある集計データは、副本文を含むデータである。宣命の場合、副本文に漢語の形態論情報を持つ語が多いので、上記語数表、語彙表のデータにおける漢語の割合が高い。そのため語数表、語彙表の取り扱いには注意が必要である。

8. おわりに

本コーパスでは、独自の漢字仮名交じりテキスト

トを作成することで、形態素解析を行うことを可能にした。その際に、表記と読みによる単位認定の問題については、同一箇所複数の形態論情報を付与することを提案し、以下のことが可能になった。

- ①主本文と副本文を交換することで、表記上分割できない短単位に複数短単位の情報を付与し、複数の情報から検索できる。
- ②宣命に特異な訓読がなされている、または他のサブコーパスと統一しにくい語について、音読の情報を付与する。
- ③複数の形態論情報を組み立て、短単位と長単位の情報を一致させる。

形態論情報の多重化は、本来は掛詞や洒落などに対応するための機能であるが、宣命はこの機能を活用することで、表記と読みの対応できないことを解消させることができた。本手法は、他の和漢混淆文などの表記と短単位の境界が合わない資料のコーパスの構築にも応用可能なものであると考えられる。

付記 本研究は国立国語研究所共同研究プロジェクト「通時コーパスによる日本語史研究の新展開」および国立国語研究所共同研究プロジェクト「表記情報と書誌形態情報を加えた日本語歴史コーパスの精緻化」の研究成果を報告したものである。

参考文献

- [1] 池上尚：『今昔物語集（本朝部）』のデータについて，鎌倉時代編Ⅰ 概要書・規程集（2016）．
- [2] 小木曾智信：多重の読みをもつテキストのコーパス化，言語資源活用ワークショップ 2016 発表論文集（2016）．
- [3] 片山久留美，小木曾智信，上野左絵：『日本語歴史コーパス江戸時代編Ⅲ近松浄瑠璃』の公開，日本語学会 2020 年度春季大会（2020a）．
- [4] 片山久留美：『日本語歴史コーパス 江戸時代編Ⅲ近松浄瑠璃』副本文の形態論情報の概要，江戸時代編Ⅲ近松浄瑠璃概要書・規程集（2020b）．
- [5] 国立公文書館デジタルアーカイブ：『続日本紀』（請求番号：特 084-0002）
<https://www.digital.archives.go.jp/das/image/M2014102020501890601>
- [6] 国立国語研究所：『日本語歴史コーパス 奈良時代編Ⅱ 宣命』（短単位データ 1.0 / 長単位データ 1.0，中納言バージョン 2.5.0）
[https://pj.ninjal.ac.jp/corpus_center/chj/nara.html#senmyo\(Ver.1.0\)](https://pj.ninjal.ac.jp/corpus_center/chj/nara.html#senmyo(Ver.1.0))（2020 年 11 月 3 日確認）
- [7] 村山実和子，小木曾智信，中村壮範：形態論情報の多重化による洒落本コーパスの質的拡張，情報処理学会研究報告 Vol.2017 CH 114, No.8（2017）．
- [8] 松崎安子，小木曾智信，中村壮範：『日本語歴史コーパス 和歌集編』 Ver.1.0 の公開，日本語学会 2019 年度秋季大会予稿集（2019）．